A systems toxicology paracetamol overdose framework – accounting for high risk individuals

Chantelle L. Mason[a], Joseph Leedale[b,1], Sotiris Tasoulis[c], Ian Jarman[a], Steven D. Webb[a,b]

[a] Department of Applied Mathematics, Liverpool John Moores University, James Parsons Building, Byrom Street, Liverpool, L3 3AF, UK.

[b] EPSRC Liverpool Centre for Mathematics in Healthcare, Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, UK.

[c] Department of Computer Science and Biomedical Informatics, University of Thessaly, Papassiopouli 2-4, GR-35100, Lamia, Greece.

[1]To whom correspondence should be addressed at Department of Mathematical Sciences, University of Liverpool, Peach Street, Liverpool, L69 7ZL, UK. Tel: +44 151 794 4049. E-mail: j.leedale@liverpool.ac.uk.

15 **Abstract**

16 The most commonly prescribed painkiller worldwide, paracetamol (acetaminophen, APAP) is
17 also the predominant cause of acute liver failure (ALF), and therefore paracetamol-induced
18 liver toxicity remains an important clinical problem. The standard clinical treatment framework
19 for paracetamol overdose currently allows for antidote therapy decisions to be made based
20 on a nomogram treatment line. This treatment threshold is lowered for patients adjudged to
21 be highly susceptible to liver injury due to risk factors such as anorexia nervosa or bulimia.
22 Additionally, both the original and adjusted clinical frameworks are highly dependent on
23 knowledge from the patient regarding time since ingestion and initial dose amount, both of
24 which are often highly unpredictable. We have recently developed a pre-clinical framework for
25 predicting time since ingestion, initial dose amount and subsequent probability of liver injury
26 based on novel biomarker concentrations. Here, we use identifiability analysis as a tool to
27 increase confidence in our model parameter estimates and extend the framework to make
28 predictions for both healthy and high-risk populations. Through pharmacokinetic-
29 pharmacodynamic model refinement, we identify thresholds that determine whether necrosis
30 or apoptosis is the dominant form of cell death, which can be essential for effective ALF
31 interventions. Using a single blood test, rather than the multiple tests required in the current
32 clinical frameworks, our model provides overdose identification information applicable for
33 healthy and high-risk individuals as well as quantitative measures of estimated liver injury
34 probability.

35 **Keywords**

36 APAP; DILI; biomarkers; *in-silico*; pharmacokinetics; identifiability.

1 Introduction

Paracetamol (acetaminophen, APAP), the most commonly prescribed painkiller in the world [1], is also the leading cause of acute liver failure (ALF) [2] and therefore represents a concerning global health issue [3]. In England and Wales, APAP poisoning results in approximately 40,000 hospital admissions, 20 liver transplants and 200 deaths per year [4]. Between 2015-2016, there was an 11% increase in deaths involving APAP in the UK [5]. In the USA, ALF is responsible for approximately 56,000 emergency room visits, 2,600 hospitalisations and 500 deaths per year [6]. Considering this, the clinical overdose intervention treatment framework is found to be surprisingly sub-optimal. Administration decisions regarding the overdose antidote, N-Acetylcysteine (NAC), are currently based upon the nomogram treatment line [7] which, though influenced by a measurement of alanine aminotransferase (ALT), is also heavily dependent on patients' knowledge of initial dose amount and time elapsed since ingestion; information which is often unreliable. Given this uncertainty, decisions on whether or not to administer the NAC antidote can be imprecise. Furthermore, unnecessary NAC administration can cause a range of side effects such as nausea, vomiting and anaphylactoid reaction, thereby exacerbating the problem of ill-informed treatment [8]. Such inaccurate treatment decisions have led to an estimated cost of £8.3 million per year in the UK since 2012 [9].

APAP is predominantly metabolised in the liver via glucuronidation and sulphation pathways, with a small fraction being oxidised into the toxic metabolite, NAPQI. Detoxification of NAPQI occurs via conjugation with hepatic stores of glutathione (GSH) [10]. Therefore, although initial dose and time since ingestion are known to be the most important indicators of overdose severity level, additional factors affecting an individual's ability to synthesise or maintain sufficiently high levels of GSH should also be considered [11]. Such factors may include age, pre-existing liver disease, concurrent use of alcohol and/or other liver-metabolised medications, genetic predispositions and acuity/chronicity of APAP use [3]. Patients with

known high-risk factors are currently measured against an amended nomogram treatment line when deciding whether or not to administer NAC [12].

Crucially, the availability of GSH is known to be heavily dependent on the nutritional level of a patient. Therefore, malnourished patients, or those suffering from eating disorders (particularly anorexia or bulimia nervosa) are deemed to be at a particularly high risk of developing liver injury following overdose of APAP [11]. Eating disorder admissions have risen by approximately 34% since 2006 in the UK and, with anorexia having the highest mortality rate of all psychiatric disorders (often linked to suicide) [13], it is essential that quick and accurate treatment decisions for these high-risk patients can be made so that fatalities can be avoided. In addition to an increase in those patients with eating disorders, the prescription of medications combining APAP with opioids is likely to increase the incidence of unintentional APAP overdose, particularly in the USA where there is currently an opioid addiction epidemic [14]. Furthermore, the number of alcohol-related hospital admissions in England rose by 22% in the 10-year period leading up to 2016 [15]. Since patients with opioid and alcohol-related issues are known to have a higher susceptibility to APAP toxicity, there is an increasingly urgent need to improve intervention efficacy for the growing number of patients in these high-risk groups.

Some of the most important goals of medicine are to reduce pain and prolong life [16] and scientific research has allowed for significant advancements in achieving these goals. However, there is also an increasing impetus to reduce the extent of animal testing required to conduct medical research [17]. Quantitative systems toxicology (QST) modelling comprises a useful tool to reduce and refine animal testing and is now considered as both an essential component of modern toxicity testing and a foundation for individualised therapeutic treatment [18]. However, the utility of a QST model, as an abstract representation of the true biology, is limited by its simplifying assumptions and consequently there are often multiple aspects of the model that contain uncertainty. Not all states of a dynamic model can be directly measured

experimentally, and conversely, not all experimental data may be useful for model calibration, since the data itself may contain errors not accounted for by the model. These limitations can raise scepticism around QST model predictions and while it would be unrealistic to attempt to completely eradicate every level of error, it is crucial that any parameter uncertainties should be assessed, reported and minimised in order for these models to be truly useful in their predictions [19]. There are many existing and developing techniques to quantify uncertainty, and the chosen method often depends on the aims of the model. Identifiability analysis can be employed to determine whether model parameters can be uniquely identified based upon the structure of the model and data used, and sensitivity analyses can provide quantification of the dependency of model outputs to perturbations in the model parameters.

We have recently developed a framework that uses a single sample, rather than multiple sample, approach to biomarker quantification in order to predict the probability of liver injury [20]. This model was optimised against fed mouse data and therefore is limited to applications relating to individuals with unimpaired clearance capacity. Through the application of uncertainty quantification techniques, we here identify areas within the original model structure that require improvement and use this knowledge to make the structure more relevant to the APAP toxicity clinical environment. There are many other *in-silico* models which focus on describing and understanding APAP-induced toxicity. Howell et al. [21] combined a large-scale, mechanistic mathematical model (DILIsym®) with *in-vitro* data to compare DILI responses across species. Whilst their model has potential utility for *in-vitro* to *in-vivo* extrapolation, parameter identifiability was not assessed within their study. Reith et al. [22] clarified the role of glucuronidation and sulphation pathways in the hepatic metabolism of APAP. Ochoa et al. [23] reported a physiologically-based pharmacokinetic (PBPK) modelling approach to predict APAP and toxic metabolite concentrations, which were then used to estimate spatiotemporal cell integrity and the elimination rates of various substances. Zurlinden et al. [24] used a Bayesian inference approach within a PBPK model to estimate

initial APAP dose and quantify its uncertainty based on conventional biomarkers. Our model aims to optimise patient stratification and overdose treatment by incorporating a panel of novel mechanism-based biomarkers of increased sensitivity [25]. Remien et al. [26] also used conventional biomarkers to predict initial dose and time since overdose. Our work in [20] extended this previous work with the inclusion of novel biomarkers HMGB1 and K18. However, there is currently only one *in-silico* APAP model which takes into consideration individuals that may have depleted GSH stores. Navid et al. (2013) developed a multi-compartmental PBPK model of APAP metabolism in order to understand how nutritional deficiencies and certain lifestyle choices, such as alcohol consumption, affect GSH regeneration [27].

PBPK models are increasingly being utilised to account for blood flow between organs, tissue partitioning, and predicting localised drug concentrations at the site of action [24, 28]. However, these large-scale models, while more physiologically relevant, favour complexity at the expense of mathematical tractability and subsequently contain an increased number of parameters that require estimation. The large scale of the models and amount of parameters to be optimised, often outweighing the amount and quality of data available for such a task, renders identifiability of model parameters problematic. Since the distribution and metabolism of APAP is well known, we opt for a smaller scale model which focuses on liver toxicity biomarkers. The smaller scale model allows for closer scrutiny of the model itself via techniques such as sensitivity analysis and identifiability analysis, providing mechanistic insight into structure, parameterisation and interactions within the system.

In this study, we extend our original APAP overdose framework [20] to quantify the effects of various factors that impact upon GSH availability to develop an improved overdose identification framework. The mathematical model is refined using techniques from uncertainty analysis to account for mechanistic changes indicative of suppressed GSH capacity and optimised against additional data to improve its scope and predictive potential. Analysis of this

improved model provides predictions of initial dose, time since ingestion and probability of liver injury for both healthy and high-risk populations.

2 Methods

2.1 Experimental data description

The mathematical model is parameterised against multiple experimental datasets. For the APAP PK element of the model, four datasets from two separate published studies [25, 26] recording APAP concentration over time in mice following intraperitoneal administration of 50, 150, 500 and 530 mg/kg doses are used in both the original and extended framework. For the biomarker PD element of the model, in the original framework one experimental dataset is used [30] which records biomarkers (GSH, ALT, HMGB1, full and fragmented K18) over time following a 530 mg/kg APAP dose. In the extended framework, this dataset is also used, but with the addition of two other datasets from two separate studies [20, 27]. The first, [20], provides dose response data for mouse biomarker concentrations GSH, ALT, HMGB1 and fragmented K18 at 5 hours following APAP doses [0,150,300,530,750,1000] mg/kg. The second, [31], provides biomarker concentrations GSH, ALT, HMGB1, full and fragmented K18 at 5 and 24 hours for both fed and fasted mice following a 530mg/kg APAP dose. All datasets described were used for optimisation of the model parameters. The data consists of a wide range of dosing scenarios to the extent that we do not extrapolate beyond these calibration ranges in subsequent model simulations.

2.2 Parameter optimisation

A multi-start technique is applied during parameter optimisation in an attempt to find the globally optimal parameter set following the restructuration of the original model [20]. All dynamic models are optimised by minimising the sum of squared errors (SSE) between model simulated output and the experimental data described in section 2.1 (fminsearch, Matlab [32]). The Matlab minimisation function uses a Nelder-Mead search algorithm to iteratively search

the parameter space until a local minimum is found [33]. Latin hypercube sampling was used to generate 1000 different initial estimates for each parameter (ranges defined in the supplementary material).

2.3 Identifiability analysis

Identifiability analysis is performed to visualise changes in SSE (deviation between model output and experimental data) following parameter perturbations and, subsequently, to determine the identifiability of each parameter in the model. We apply a method of identifiability analysis similar to the profile likelihood approach defined by the FRIAS research group [34]. Parameter estimates are either identifiable (unique minimum), practically unidentifiable (monotonic response), or structurally unidentifiable (negligible response) [35]. During this analysis, each individual parameter is tested separately for identifiability. This “test parameter” is varied by 20% intervals (from -50% to +200% of its original value). In each iteration, the modified test parameter is fixed, while all the other parameters in the model are varied in two ways: fixed at the optimum values (as found from previous multi-start optimisation); or randomly sampled from a Latin hypercube (bounds for sampling can be found in the supplementary material). For each test parameter iteration, the parameter set corresponding to the lowest SSE value is then identified. In this analysis, an identifiable parameter is defined as a parameter which, when perturbed from its initial (optimal) value (both positively and negatively), results in an increased SSE and therefore predicts a greater error between the model output and the data. If the SSE increases on only one side (i.e. in the positive or negative direction) of the test parameter, and remains relatively unchanged on the other side, this parameter is defined as practically unidentifiable; that is, either increasing or decreasing the test parameter value causes an increased error between the model output and the data. However, since the error between model output reduces as we head towards the test parameter, but then remains relatively unchanged for further perturbations in that direction, we cannot be confident that the parameter is uniquely optimal, since there are multiple values

that provide the same approximate SSE. However, since the optimum path does change in one direction, we have confidence that we are capturing some of the structure, and often these kind of 'practical unidentifiabilities' can be resolved by including more calibration data [34]. If the SSE value does not change either side of the original test parameter value, this parameter is known as structurally unidentifiable. This means that parameter optimisation via data-fitting is relatively insensitive to the choice of this parameter; the parameter cannot be uniquely determined and therefore even if removed entirely, values of other parameters could be suitably adjusted to compensate for the change in the model structure. During both sensitivity and identifiability analyses, model parameters are perturbed and subsequent changes in model output are studied. However, parameters are not re-optimised during sensitivity analysis. Identifiability analysis seeks to determine whether distinct model parameterisations could provide the same model solution.

2.4 Model refinement

2.4.1 APAP pharmacokinetic model

Three datasets from three separate published studies [20, 26, 27] were used to parameterise the APAP pharmacokinetic (PK) model which is developed as an extension to our previous framework [20]. The PK model structure remains unchanged and is summarised below for completeness. Note that all the data used was obtained in studies of APAP dosing in mice (fed or fasted). Two ordinary differential equations (ODEs) are used to represent changes in APAP concentration within two PK compartments (central and peripheral) in the following system,

$$\frac{d[C_c]}{dt} = \frac{k_a D_0 e^{-k_a t}}{V_c} + k_{21}[C_p]\frac{V_p}{V_c} - k_{12}[C_c] - k_{el}[C_c], \quad (1)$$

$$\frac{d[C_p]}{dt} = k_{12}[C_c]\frac{V_c}{V_p} - k_{21}[C_p], \quad (2)$$

where $[C_c]$ is the central compartment concentration (μmol/l) of APAP; $[C_p]$ is the peripheral compartment concentration (μmol/l) of APAP; and k_a represents the absorption rate (h^{-1}) from where APAP is administered (the peritoneal cavity in this case). The initial dose (μmol/kg) is given by D_0; k_{21} represents the transfer rate (h^{-1}) from peripheral to central compartment; k_{12} represents the transfer rate (h^{-1}) from central to peripheral compartment; V_p is the theoretical volume (l/kg) of the peripheral compartment; V_c is the theoretical volume (l/kg) of the central compartment; k_{el} represents the overall elimination rate (summation of excretion and metabolism processes) (h^{-1}); and t represents the time variable (h). Further details of the APAP pharmacokinetic model can be found in our previous publication [20].

2.4.2 Pharmacodynamic models

Previously, the pharmacodynamic (PD) element of the model was parameterised using a dataset consisting of GSH and biomarker (ALT, HMGB1, full and fragmented K18) time-course concentrations following a 530 mg/kg intraperitoneal APAP dose [20]. In this paper, we extend this optimisation to also include dose-response data consisting of plasma biomarker concentrations at 5 hours following APAP doses ranging from 0-1000 mg/kg [20] and a dataset consisting of biomarker concentrations at 5 and 24 hours for both fed and fasted mice following a 530 mg/kg APAP dose [31]. This extension is necessary in order to account for differing mechanisms of cell death, i.e. apoptosis versus necrosis, and also to account for an increased dosing range more representative of the clinical environment.

2.4.2.1 Glutathione dynamics

In our model, paracetamol toxicity biomarker dynamics are assumed to be directly dependent on GSH depletion; i.e., during APAP overdose cases when GSH pools deplete, NAPQI accumulates potentially leading to liver toxicity and associated biomarker release. The GSH model component remains identical to our previously defined model [20], namely:

$$\frac{d[gsh]}{dt} = k_o(gsh_0 - [gsh]) - \frac{\xi k_{el} C_c [gsh]}{[gsh] + k_{pr}}, \quad (3)$$

where k_o is the basal removal rate (h^{-1}) of GSH (including background usage); gsh_0 is the baseline concentration (μmol/l) of GSH in the APAP-free steady state; ξ is the proportion of eliminated APAP that is transformed into NAPQI; k_{el} is the APAP elimination rate; k_{pr} is the ratio of NAPQI forming other protein adducts, relative to being detoxified by GSH, and $[gsh]$ represents the concentration (μmol/l) of GSH. Further details on the derivation of the GSH model can be found in the supplementary material of our previous study [20]. For the fasted case, basal GSH dynamics are modified to simulate a delay in GSH repletion due to depleted co-factors stemming from the reduced food intake. These slower dynamics are incorporated by rescaling k_o by an additional parameter, $\delta < 1$ (so that k_o becomes $\delta k_o < k_o$).

2.4.2.2 ALT and HMGB1 dynamics

The toxic response of biomarkers ALT and HMGB1 to APAP overdose is mathematically modelled as in the previous study [20], namely:

$$\frac{d[r]}{dt} = r_0 k_{out} \left(\frac{R_{50}^n + gsh_0^n}{R_{50}^n} \right) \left(1 - \frac{g_{max}[gsh]^n}{R_{50}^n + [gsh]^n} \right) - k_{out}[r], \quad (4)$$

Where $[r]$ is the biomarker concentration (μmol/l), r_0 is the respective biomarker baseline (μmol/l); k_{out} is the natural decay rate (h^{-1}) of the biomarker; R_{50} represents the concentration (μmol/l) of (GSH) at which the biomarker response to GSH is half its maximal rate; and n is a Hill coefficient indicating the steepness of the biomarker response [36]. Parameter g_{max} represents the maximal regulatory effect of GSH on biomarker production. In the fed case, we set $g_{max} = 1$ such that GSH depletion is solely responsible for augmented biomarker production and therefore for the APAP-free steady state, the biomarker is only produced at low, basal levels maintaining the background steady state value, r_0. This element of the model is therefore identical to the previous study. However, experimental observations show that

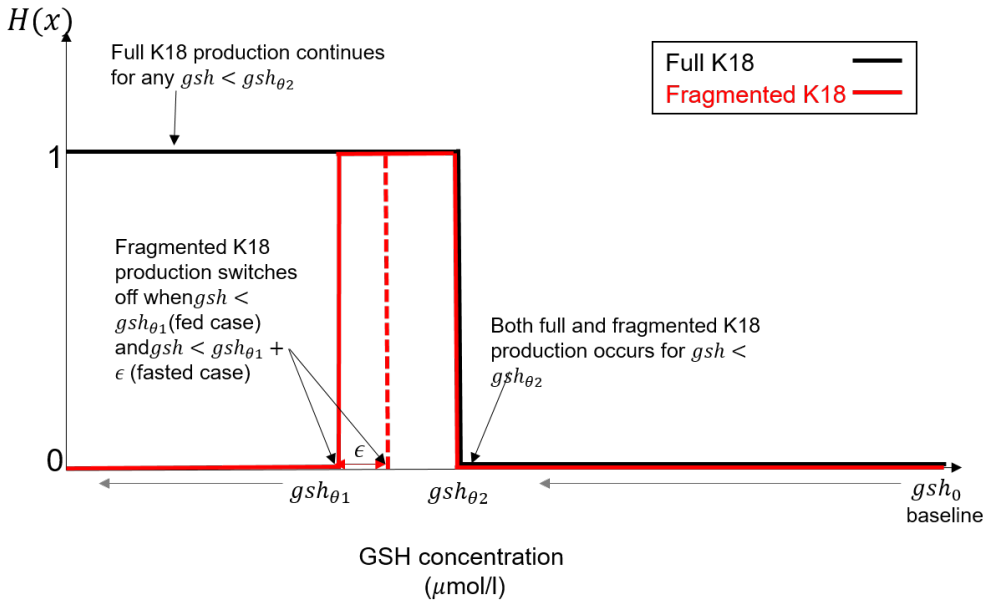
background biomarker levels are higher for fasted animals. We therefore take $g_{max} < 1$ in the fasted case to account for these higher background values.

2.4.2.3 K18 dynamics

Keratin-18 (K18) is an intermediate filament protein expressed in abundant levels in hepatocytes that undergoes caspase-mediated cleavage during apoptosis, resulting in the release of fragmented K18 upon cell death. This feature makes K18 a useful biomarker to distinguish between apoptosis and necrosis as the presence of full (as opposed to fragmented) K18 instead suggests the occurrence of necrosis [37]. Full and fragmented K18 are therefore modelled as necrotic and apoptotic forms of the same single biomarker, K18. However, we could not find sufficient data for K18 to properly parameterise a model of the form (4) for both full and fragmented K18. We therefore adopt a simple form of K18 using piecewise linear representations of (4), as illustrated in Figure 1. That is, the dynamics of the necrotic marker, full K18, are modelled in the following way,

$$\frac{d[k18]}{dt} = r_0^{18} k_{out}^{18} + k_{max} k_{in}^{18} H\left(gsh_{\theta_2} - [gsh]\right) - k_{out}^{18}[k18], \quad (5)$$

where r_0^{18} is the baseline concentration (μmol/l) of full K18; k_{out}^{18} is the natural decay rate of the biomarker; k_{in}^{18} is the production rate of the biomarker; gsh_{θ_2} is the GSH threshold below which additional K18 production is induced due to cell death; $[k18]$ represents the concentration of full K18 (μmol/l); and k_{max} is a measure of the production capability of full K18 (i.e., $0<k_{max}<1$; since there is a finite quantity of cells, there is a maximum amount of biomarker that can be produced). In the fed case, we take $k_{max} = 1$. Since fasting in mice causes extensive cell loss at early time points [31], the amount of biomarker able to be produced from a smaller amount of cells will therefore be smaller. To account for this, in the fasted case, we take $k_{max} < 1$. $H(x)$ is the Heaviside function where $H(x) = 1$ when $x > 0$ and $H(x) = 0$ when $x < 0$.



280

281 **Figure 1: Relationship between full and fragmented K18 biomarker production and GSH**
282 **concentration.** Black lines represent the full K18 relationship, and red lines represent the
283 fragmented K18 relationship (solid lines represent the fed case and dashed lines represent
284 the fasted case). As GSH depletes from baseline, gsh_0, and reaches a certain threshold,
285 gsh_{θ_2}, production of full and fragmented K18 begins. Full K18 production continues for any
286 GSH concentration below this threshold; however, fragmented K18 production ceases beyond
287 a GSH concentration of gsh_{θ_1} in the fed case and $gsh_{\theta_1} + \varepsilon$ in the fasted case.

288 Similarly, the dynamics of fragmented K18 are modelled in the following way,

$$\frac{d[fk18]}{dt} = r_0^{f18} k_{out}^{f18} + k_{in}^{f18} \left(H\left([gsh] - gsh_{\theta_1} - \varepsilon\right) - H\left([gsh] - gsh_{\theta_2}\right) \right) - k_{out}^{f18}[fk18], \quad (6)$$

289 where r_0^{f18} is the baseline concentration (μmol/l) of fragmented K18; k_{out}^{f18} is the natural decay
290 rate (h^{-1}) of the biomarker; k_{in}^{f18} is the production rate (h^{-1}) of the biomarker; gsh_{θ_2} is the GSH
291 threshold below which augmented production of fragmented K18 production is initiated,
292 and gsh_{θ_1} is the GSH threshold below which augmented production ceases due to a switch
293 from apoptosis to necrosis. $[fk18]$ is the concentration (μmol/l) of fragmented K18. For the fed
294 case, production begins at gsh_{θ_2} and ceases at gsh_{θ_1}. For the fasted case however, necrosis
295 will be more apparent than apoptosis, and less GSH depletion will be required before
296 apoptosis is no longer sustainable. To account for this, ε represents the change in GSH

297 threshold denoting the switch to necrosis, i.e., $\varepsilon = 0$ in the fed case and $\varepsilon > 0$ in the fasted
298 case.

299 All initial conditions and parameter values for the PKPD model can be found in the
300 supplementary material.

301 ***2.5 Statistical analysis***

302 *2.5.1 Virtual population simulation*

303 Model predictions are made for three virtually simulated populations: healthy, high-risk, and a
304 mixture of healthy and high-risk individuals. Healthy populations are based on biomarker
305 concentrations simulated with parameter values derived from fed data. High-risk populations
306 are based on biomarker concentrations simulated with parameter values derived from fasted
307 data. The mixed population is based on a weighting of the biomarker concentrations simulated
308 with parameter values derived from fed/fasted data, with respect to the proportion of
309 healthy/high risk patients that are seen in the clinic. Craig et al. [38] analysed overdose
310 patterns in 663 patients over 16 years and found that 42.3% patients had psychiatric history,
311 45.3% had alcohol abuse, and 44.7% combined the overdose with alcohol. In line with this,
312 44.1% of the population is assumed to be high-risk and 55.9% are assumed to be healthy in
313 our mixed population case.

314 Each virtual population dataset consists of 1000 independent and individually distributed *in-*
315 *silico* individuals, given a random dose selected from a uniform distribution of range 0-1000
316 mg/kg. Biomarker concentrations are subsequently extracted at a random time-point from a
317 uniform range between 0-24 hours. Simulated concentrations are normalised in the range [0,1]
318 using the min-max normalisation method [39] to account for varying orders of magnitude.
319 Experimental noise is replicated in the *in-silico* data by applying observed *in-vivo* standard
320 deviations in biomarker concentrations from an APAP study performed by Antoine et al. (2009)
321 (ALT s.d = 11.22, HMGB1 s.d = 0.00097, K18 s.d = 2.39, fragmented K18 s.d = 0.12 μmol/l).

2.5.2 Predicting time since administration and initial dose

2.5.2.1 Multiple linear regression

Normality tests indicate that the simulated data is non-normally distributed [40]. Improvements to the linear model, made by employing interaction/polynomic terms, are insignificant, and hence a robust multiple linear regression model [41] is applied to the *in-silico*-derived data in order to predict time since administration and initial dose.

2.5.2.2 Visualisation

For each *in-silico* individual, the t-Distributed Stochastic Neighbour Embedding (T-SNE) method [42] is applied to visualise the dataset constituted by the aforementioned variables (APAP, GSH, ALT, HMGB1, K18, fragmented K18). Two-dimensional scatter plots of the embedded data are employed in order to examine class structure and separability whilst retaining model variation.

2.5.2.3 Classification

In order to test the predictive potential of biomarker concentrations, critical ranges for predicting time since administration are defined as (0-2], (2-5], (5-10], (10-15] and (15-24] hours. For dose, the ranges are [0-200], [201-400] and [401-1000] mg/kg, capturing therapeutic, small, and large (overdoses), respectively. A number of classification techniques are applied and compared, further details of which can be found in the supplementary information of our previous study [20].

2.5.3 Predicting probability of liver injury

Our previously defined binary logistic regression framework [20] uses experimental biomarker time-course data [30] to predict a corresponding histology score for each mouse. The whole panel of biomarkers was tested, and analysis found HMGB1 concentration to be the most

significant in predicting the probability of liver injury. The resultant logistic regression model is used here in combination with PK-PD model simulations to predict the drug-induced liver injury (DILI) probability [43].

3 Results

3.1 Identifiability analysis – original model structure

Parameters in the original model structure are individually perturbed to visualise resultant differences between model output and experimental data. These changes are assessed to determine the identifiability of each parameter and identify where changes may be necessary. As seen in Figure 2, 10 out of 21 parameters are identifiable. The ALT component of the model is identifiable, and all but one of the parameters in the PK component are identifiable. However, the HMGB1 component of the model is structurally unidentifiable, and practical unidentifiabilities exist in all other elements of the model. Five parameters are identified as practically unidentifiable. There are 3 within the GSH component: the ratio of NAPQI forming other protein adducts relative to being detoxified by GSH (k_{pr}); the proportion of eliminated APAP that is transformed into NAPQI (ξ); the basal removal rate of GSH (k_o). The remaining practical unidentifiabilities are found within the K18 and fragmented K18 components of the model: the decay rate of full K18 ($K18_{kout}$); and the fragmented K18 hill coefficient ($f18_n$). The remaining 6 parameters are structurally unidentifiable. These include the theoretical volume of the peripheral compartment (V_p) from the PK component and all parameters from the HMGB1 component: the hill coefficient (H_n); the GSH concentration at which the augmented production rate of HMGB1 reaches 50% of its maximum (H_{IC50}); and the decay rate (H_{kout}). Two out of three parameters within the fragmented K18 component are structurally unidentifiable: the GSH concentration at which the augmented production rate of fragmented K18 reaches 50% of its maximum ($f18_{IC50}$); and the decay rate ($f18_{kout}$).

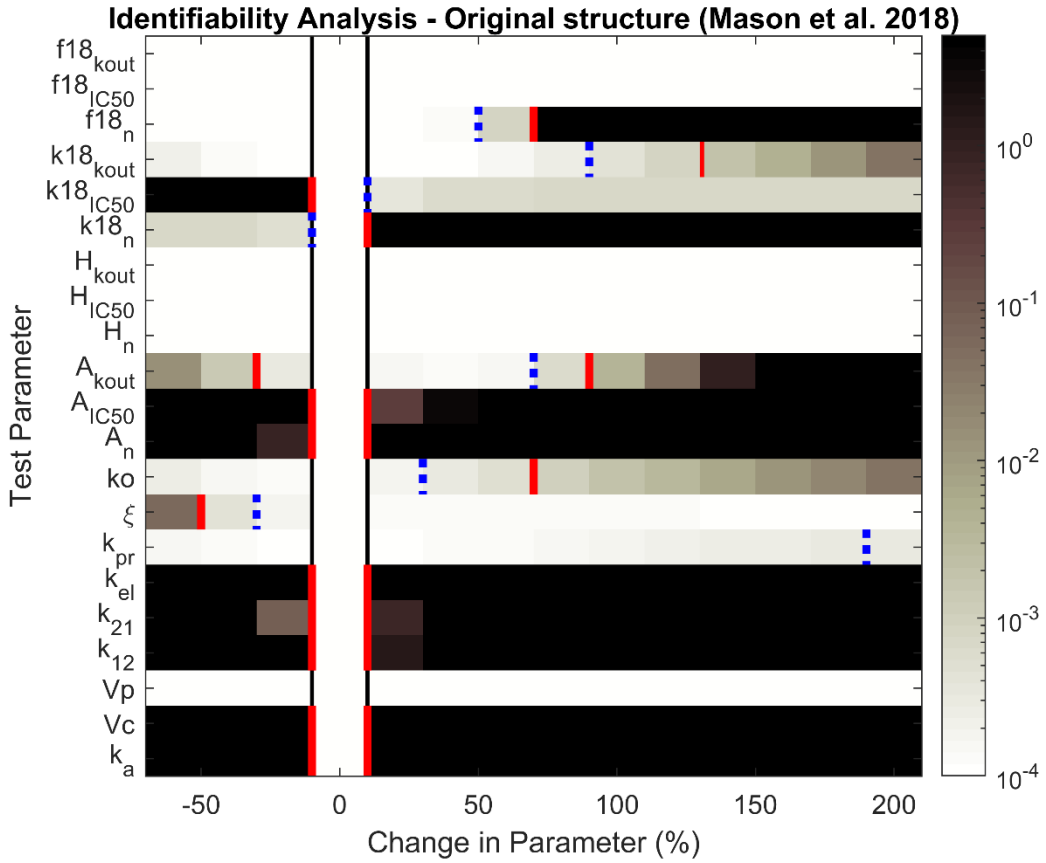


Figure 2: Identifiability analysis of the original model structure [20]. Each test parameter, was fixed at 20% intervals, and the other parameters in the model allowed to vary. The percentage change of resultant optimised function values (SSE) are plotted at each interval (on log-scale). The lowest SSE change is represented by a white box; darker colours show an increase in SSE change. The highest SSE change is represented by a black box. Blue dashed bounds indicate where the parameter is identifiable at the 0.5% level. Red bounds indicate where the parameter is identifiable at the 1% level. A parameter is identifiable at the 1% level if it is bounded by red in both the positive and negative parameter-change directions. If the parameter is bounded by red in one direction and blue in the other direction, it is identifiable at the 0.5% level but practically unidentifiable at the 1% level. A parameter is practically unidentifiable if it is bounded by red/blue in either the positive or negative parameter change direction, and unbounded in the opposite direction. A parameter is structurally unidentifiable if it is unbounded in both positive and negative parameter-change directions (no red/black bounds exist for the whole range of parameter changes). For example, A_{IC50} is identifiable at the 1% level, $K18_{IC50}$ is identifiable at the 0.5% level but practically unidentifiable at the 1% level, $f18_n$ is practically unidentifiable, V_p is structurally unidentifiable.

3.2 Re-parameterisation

Following model restructuration and re-parameterisation against the multiple datasets (increased dose range and fasted data), we identify a number of parameter adjustments. In

cases where parameters are common to both model structures, percentage changes in their values following re-parameterisation are shown in Table 1. The baseline level of GSH, GSH_0, is now optimised against the data rather than fixed as in the previous model. The resultant value reduces by almost 20% to 559.47 μmol/l. The proportion of CYP-activated APAP that is transformed into NAPQI, ξ, increases to around 80%, but there is also an increased level of NAPQI detoxification resulting from the re-parameterisation (represented by a 22.1% decrease in parameter k_{pr}, the ratio of NAPQI forming other protein adducts relative to detoxification). The new parameter, δ, incorporates an effective delay in GSH repletion in the fasted case due to depleted co-factors. The optimised value is 0.0483 and considerably reduces the timescale of GSH dynamics in the fasted case.

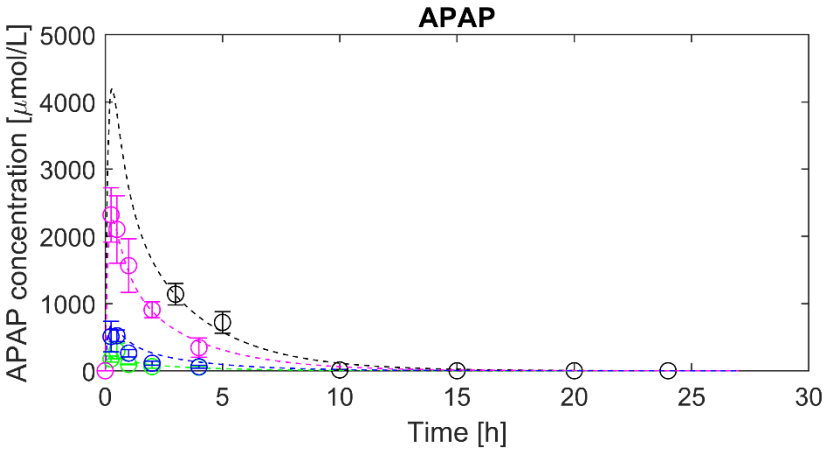
		Optimised value		
Model	**Parameter**	**Original**	**Re-parameterised**	**% change**
APAP	k_a (h^{-1})	9.05	8.6152	**-4.8%**
	k_{el} (h^{-1})	0.52	0.5459	**+5%**
	k_{12} (h^{-1})	0.42	0.4502	**+7.2%**
	V_c (l/kg)	0.02	0.0220	**+10%**
	V_P (l/kg)	0.01	0.2102	**+2000%**
	k_{21} (h^{-1})	1.01	1.0315	**+2.1%**
GSH	gsh_0 (μmol/l)	696.9136	559.47497	**-19.7%**
	k_{pr}	71.06	55.33401	**-22.1%**
	ξ	0.68	0.80571	**+18.5%**
	k_o (h^{-1})	0.25	0.78807	**+215%**
ALT	R_0 (μmol/l)	0.7626	0.7626	**0%**
	n	9.26	4.3324	**-53.2%**
	R_{50} (μmol/l)	227.67	35.6531	**-84.3%**
	k_{out} (h^{-1})	0.0002	0.0004	**+1%**
HMGB1	R_0 (μmol/l)	0.0005	0.0005	**0%**
	n	4.90	2.3445	**-51.2%**
	R_{50} (μmol/l)	399.08	75.2828	**-81.1%**
	k_{out} (h^{-1})	0.35	0.0964	**-72.5%**
Full K18	r_0^{18}(μmol/l)	0.0146	0.0088	**-39.73%**
	k_{out}^{18} (h^{-1})	0.0007	0.0031	**+342.9%**
Fragmented-K18	r_0^{f18}(μmol/l)	0.0642	0.0977	**+52.2%**
	k_{out}^{f18}(h^{-1})	0.02	0.0031	**-84.5%**

Table 1: Parameter changes following model refinement. Any parameter that remained within the new model structure is defined, with its original value and the re-parameterised value. Percentage changes are also defined.

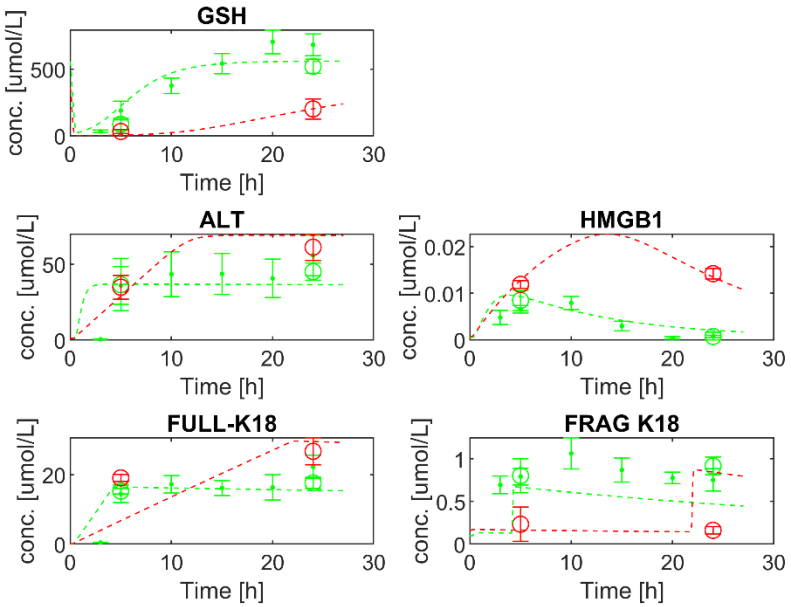
Following re-parameterisation, we found that an increased amount of GSH depletion is required for the GSH-induced ALT response to be half of its maximal value, i.e., 67% GSH

404 depletion was required before whereas almost 94% is required now. Although more GSH
405 depletion is also required for HMGB1 induction, this biomarker response is still faster than that
406 of ALT, since it has reached 50% maximal response rate at around 86.5% GSH depletion. For
407 Full-K18, results from the optimisation suggest that augmented production of the biomarker
408 will occur when GSH decreases below 174.5205 μmol/l (~68.8% depletion). Fragmented K18
409 is also induced at this level of GSH; however, if GSH is further depleted to 167.3663 μmol/l,
410 augmented production of this apoptosis marker would cease, and the necrotic, full version of
411 the biomarker would then dominate. In the fasted case, this threshold increases by 5.0457
412 μmol/l (ε) such that the necrotic switch (as indicated by the absence of K18 fragmentation)
413 occurs when GSH is depleted beyond 172.412 μmol/l. Optimised model simulations are
414 plotted and compared with the time-course data (Figure 3) and the dose-response data (Figure
415 4). Dose-response data was unavailable for full K18 at the time of investigation. Therefore,
416 the K18 parameters within the model structure are predominantly optimised against high-dose
417 data which may explain the slight under-prediction shown in the fragmented K18 simulations.

A

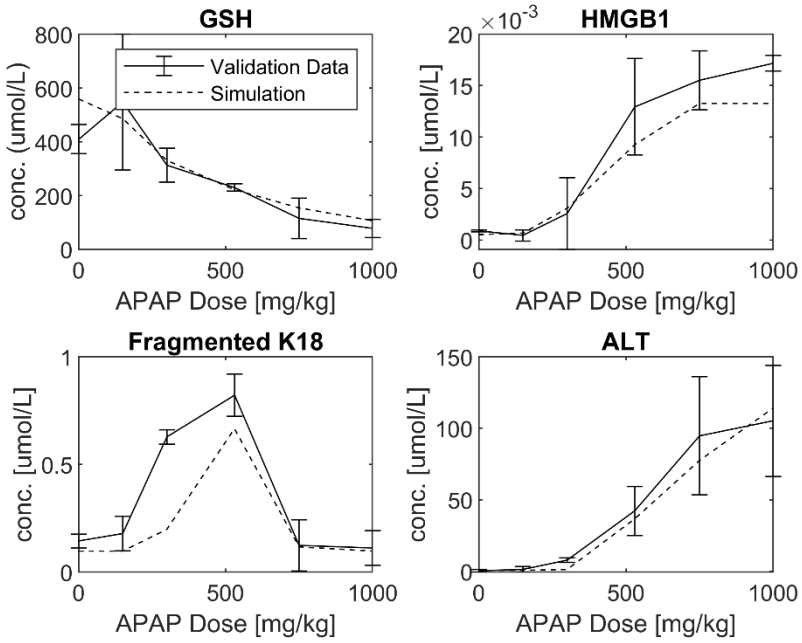


B



418 **Figure 3: Model simulations versus time-course data.** (A): APAP concentration simulations
419 are plotted for 50 (green), 150 (blue), 500 (pink) and 530 (black) mg/kg doses of APAP. (B):
420 Fed and fasted simulations (green and red respectively) are plotted for GSH and biomarkers
421 (ALT, HMGB1, Full and fragmented K18) following a 530 mg/kg dose of APAP.

422



423

424 **Figure 4: Model simulations versus dose-response data.** Dashed lines represent dose-
425 response simulations of GSH and biomarkers (ALT, HMGB1 and fragmented K18) in fed mice
426 for a range of APAP doses (0,150,300,530,750,1000 mg/kg). Solid lines represent the
427 experimental data each model was calibrated against.

428

429 ***3.3 Identifiability analysis – re-parameterised model***

430 Parameters in the refined model are individually perturbed to visualise resultant differences
431 between model output and experimental data. These changes are assessed to determine the
432 identifiability of each parameter. 16 out of 27 parameters are now identifiable. Parameter
433 identifiability has remained for the ALT component, and the GSH component of the model is
434 now also completely identifiable. The PK component remains identifiable, other than the
435 volume of the peripheral compartment (V_p), which remains structurally unidentifiable. There
436 are some unidentifiabilities still present within this updated model. Other structurally
437 unidentifiable parameters include: H_{IC50}, H_{kout} and H_{gmax} from the HMGB1 component; and
438 $K18_{kout}$, f_{kin}, f_{kout}, gsh_{θ_1} and ε from the full and fragmented K18 component. Practical

unidentifiabilities remain for two parameters: the HMGB1 hill coefficient (H_n) and the production capability of full K18 (k_{max}).

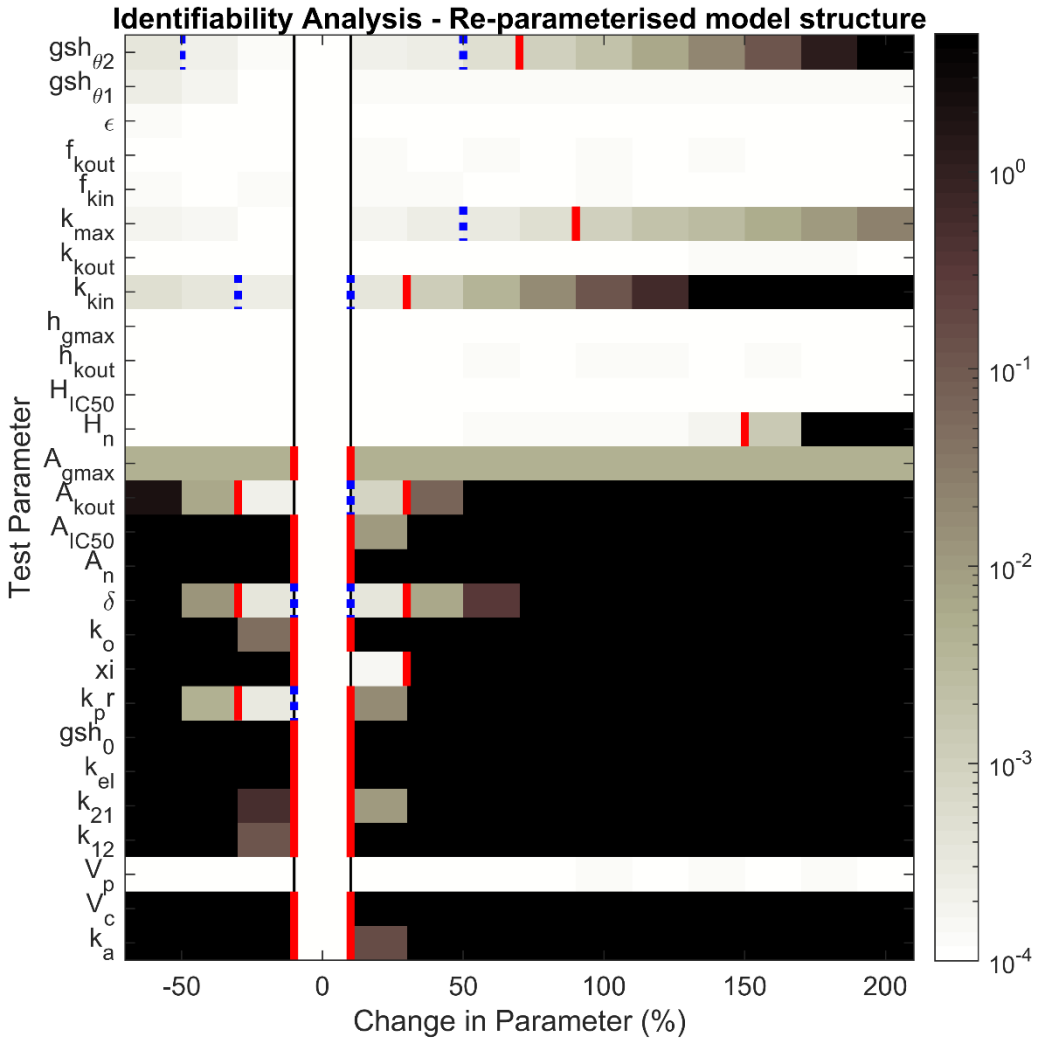
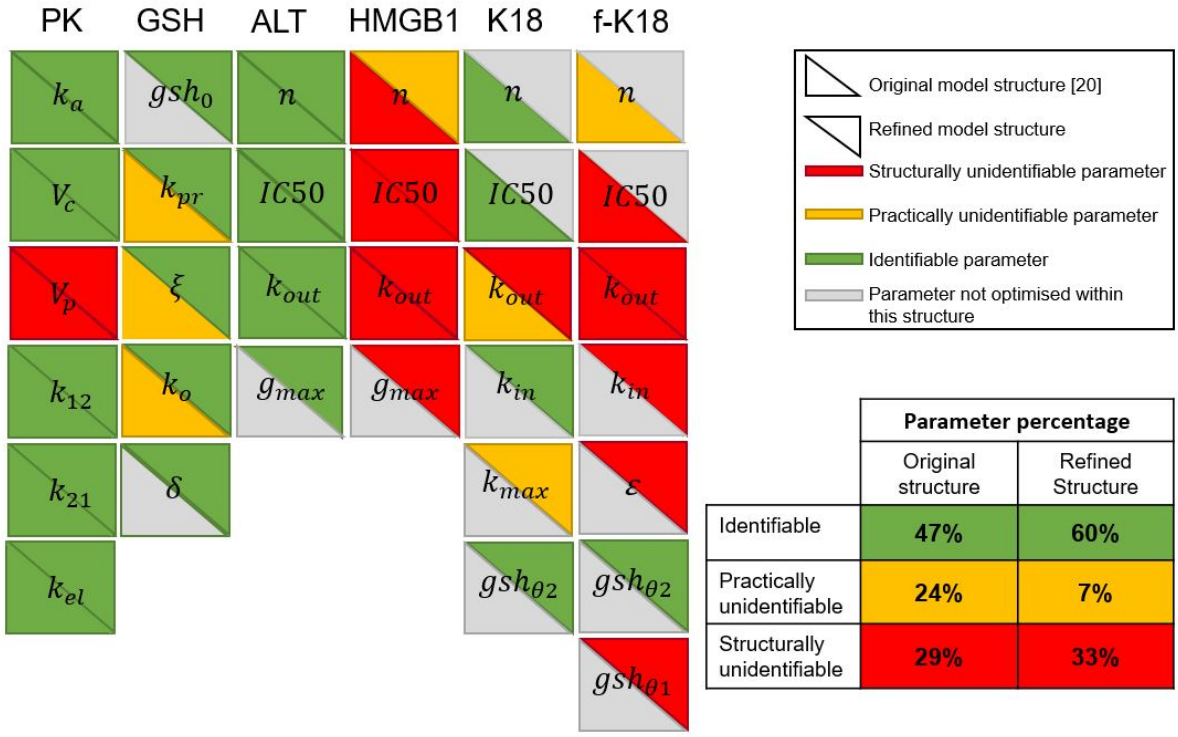


Figure 5: Identifiability analysis of the re-parameterised model structure. Figure annotation is the same as for Figure 2**.** Example results from figure: A_{IC50} is identifiable at the 1% level, H_n is practically unidentifiable, V_p is structurally unidentifiable.

Less than half of the parameters (47%) are identifiable in the original model structure (Figure 6). Only the ALT component of the model is completely identifiable. 24% of the model parameters are practically unidentifiable and 29% are structurally unidentifiable. However, following model refinement, the identifiability results have improved; the percentage of identifiable parameters has increased to above half of the parameters (60%). Parameter

identifiability has been maintained for the ALT component of the model and the percentage of practical unidentifiabilities has reduced from 24% to 7%. The percentage of structural unidentifiabilities is approximately the same (29% compared to 33%). Whilst the GSH component is unidentifiable in the original model, it is now completely identifiable in the refined model. Although unidentifiabilities have improved for parameters in the HMGB1, K18 and f-K18 components, there are still some that remain.



	Parameter percentage	
	Original structure	Refined Structure
Identifiable	47%	60%
Practically unidentifiable	24%	7%
Structurally unidentifiable	29%	33%

Figure 6: **Parameter identifiability comparison between the original and refined model.** Each column represents a different sub-component of the APAP PKPD model. Each element (or square) represents the parameter's identifiability within the model. Triangles in the bottom left-hand corner of each element represent the parameter's identifiability in the original model structure. Triangles in the top right-hand corner of each element represent the parameter's identifiability in the refined model structure. If the triangle is red, the parameter is structurally unidentifiable; if the triangle is yellow, the parameter is practically unidentifiable; if the triangle is green, the parameter is identifiable. If the triangle is grey, the parameter was not present within the corresponding model structure, and therefore could not be tested for identifiability.

3.4 Predicting an exact time since administration of initial APAP dose

Simulations of the refined model create *in-silico* derived populations. Robust multiple linear regression is applied to biomarker concentrations from each population in order to predict an exact time since ingestion and initial APAP dose. For the healthy population, an exact time since administration can be estimated with an error of approximately 3.9 hours, and an exact initial dose amount can be estimated with an error of approximately 66.14 mg/kg (Table 2).

	Healthy Population		High-risk Population		Mixed Population	
Predictors	**Time**	**Dose**	**Time**	**Dose**	**Time**	**Dose**
APAP Conc.	-16.549*** (1.1785)	673.985*** (17.0745)	-9.5093*** (0.5963)	1295.366*** (47.8525)	-18.122*** (0.9542)	909.341*** (17.8005)
ALT conc.	8.972*** (0.9533)	515.625*** (13.8110)	5.5034*** (0.6537)	695.2118*** (52.4551)	1.4573 (0.9228)	655.0837*** (17.2145)
HMGB1 conc.	-22.035*** (0.9113)	-23.3880 (13.2028)	-8.8119*** (0.5440)	367.0774*** (43.6578)	-17.869*** (0.7705)	289.2333*** (14.3736)
FullK18 conc.	7.521*** (1.0632)	479.137*** (15.4035)	21.0427*** (0.4743)	-154.2403*** (38.0652)	22.3833*** (0.9145)	229.0486*** (17.0605)
Fragmented K18 conc.	-4.599* (1.3895)	151.975*** (20.1309)	3.8472*** (0.5962)	86.7133 (47.8445)	-0.1079 (0.8321)	33.8383* (15.5231)
Constant	14.918*** (0.3074)	77.247*** (4.4529)	5.8083*** (0.2527)	-80.145*** (20.2759)	12.6308*** (0.3694)	-65.6354*** (6.8914)
Residual Std. Error (df == 994)	3.907	66.14	2.118	184.1	3.485	73.73
Note: *p<0.05; **p<0.01; ***p<0						

Table 2: Robust multiple linear regression analysis results. Independent variable coefficients for predicting time and dose respectively for healthy, high-risk and mixed populations. The first number in each element of the table represents the biomarker coefficient in the regression model and the second number represents the corresponding error. For example, -16.549 is the APAP concentration coefficient in the healthy population model predicting time since administration, and this coefficient has an error of 1.1785. The standard error of overall model predictions is provided. The significance of each biomarker in the model is indicated by the number of asterisks (see note).

Time since administration can be predicted more accurately in the high-risk population, with a standard error of 2.118 hours. It is much harder to predict the initial dose in the high-risk population, however, with the standard error being almost triple that of the model for the healthy population (184.1 mg/kg). Predictions for an assumed mixed population are similar to that of the healthy population, with a slight improvement in predicting time since administration (standard error 3.485 hours) and a slight reduction in accuracy for predicting initial dose (standard error 73.73 mg/kg). All biomarkers are significant in predicting time since administration in a healthy population and high-risk population independently. However, when assuming a mixed population, ALT and fragmented K18 are no longer significant. When predicting dose in a mixed population, all biomarkers are significant; however, HMGB1 is not significant in the healthy population, and fragmented K18 is not significant in the high-risk population.

3.5 Identifying time/dose category following APAP dose

T-SNE visualisation is applied to the *in-silico* derived data to investigate time/dose class structure and separability and subsequently various classification techniques are also employed, using the biomarker concentrations of the *in-silico* observations in an attempt to classify a time/dose category. Differences in levels of discrimination regarding initial dose and time since administration for healthy, high-risk and mixed populations can be seen for each case by embedding the *in-silico* derived data in 2-dimensions using T-SNE (Figure 7). Both variables, time since administration and initial dose, can be reasonably separated in all healthy, high-risk and mixed populations. This result is supported by the reasonable accuracy rates of the classification techniques. Visualising the healthy population (Figure 7A), shows that there is a cluster of observations on the right-hand-side where the time since administration is difficult to discriminate from the biomarker concentrations. Visualising this same cluster of observations with regards to the initial dose, however, shows that these correspond to low-dose situations (Figure 7B). From measuring biomarker concentrations, a

correct time since administration category could be predicted with 69.9% accuracy for the healthy population, and a correct dose category can be predicted with 91.5% accuracy. These results are indicated in Table 3. A time since administration category is easier to predict in the high-risk category (80.4% accuracy). However, predictions for initial dose are less accurate than the healthy population (79%). If an observation is assumed to be taken from a mixed population, the prediction accuracy is similar to that of a healthy population, with results of 69.94% and 89.5% for time since administration and initial dose respectively. In our previous study, a correct time category could be predicted with 72.8% accuracy and a correct dose category could be predicted with 86.5% accuracy [20]. For a healthy population, the time classification results have slightly worsened. However, the dose classification results have slightly improved.

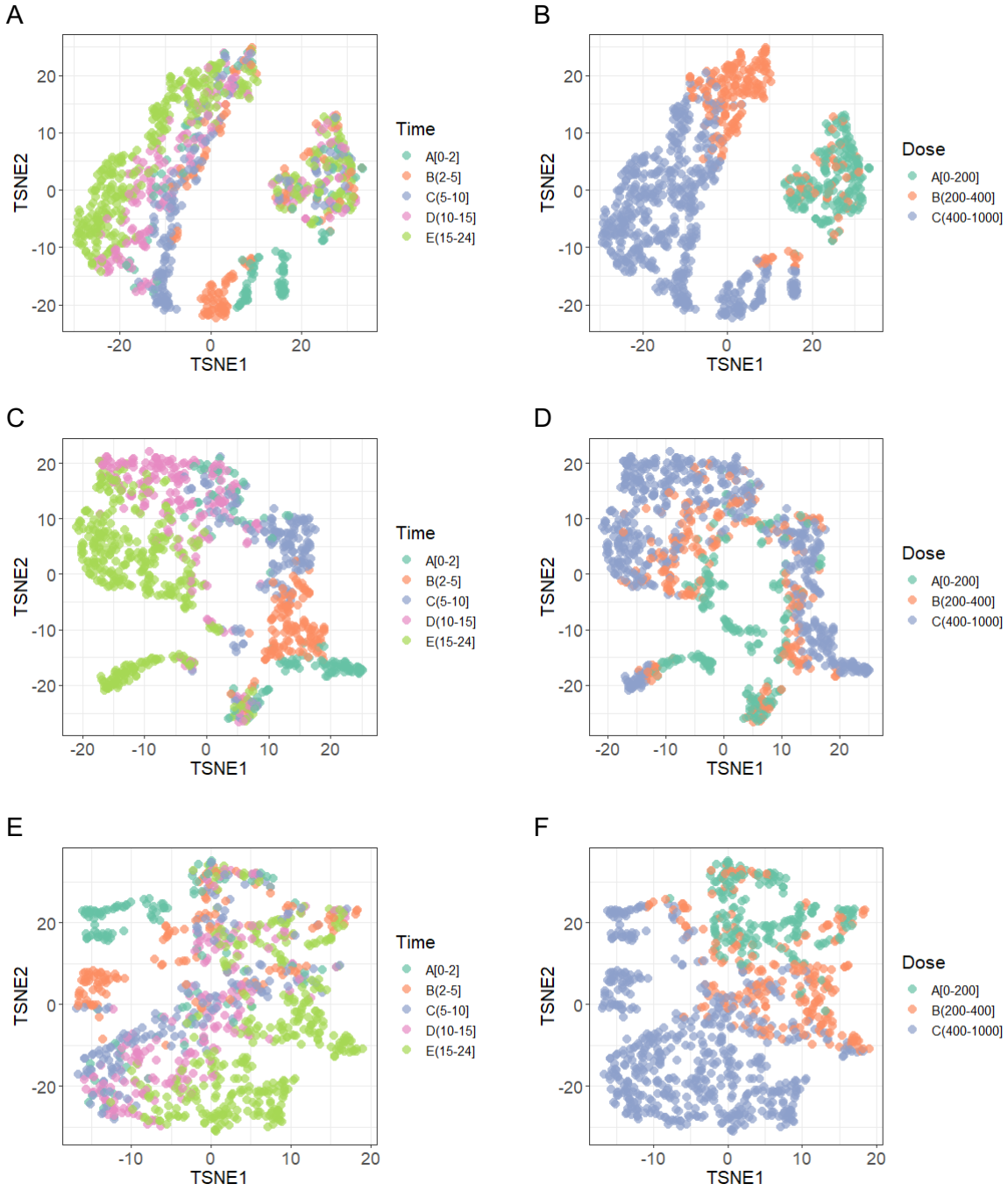


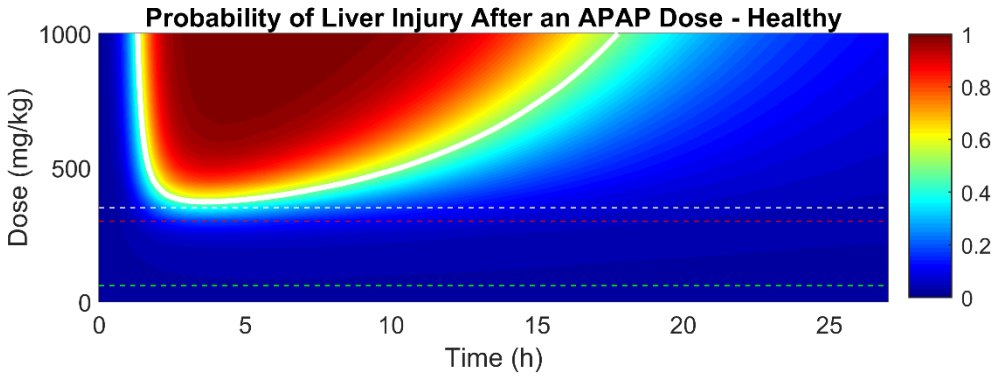
Figure 7. Visualisation and classification of time-since-administration and dose results for healthy, high-risk and mixed populations. For time-since-administration, dark green represents class [0-2), orange represents [2-5), blue represents [5-10), pink represents [10-15) and pale green represents [15-24) hours. For dose, green represents [0-200], orange represents [201-400] and blue represents [401-1000] mg/kg. TSNE visualisations of *in-silico* mouse observations with respect to time since administration and dose can be seen in (A)-(B) for the healthy population, (C)-(D) for the high-risk population and (E)-(F) for the average population.

	Healthy Population		High-risk Population		Mixed Population	
Classification Method	**Time**	**Dose**	**Time**	**Dose**	**Time**	**Dose**
KNN Regression	66.2%	91.5%	80.4%	79%	69.5%	87.4%
Naïve Bayes	64.4%	91.3%	76.2%	72.7%	68.2%	88.5%
Multinomial logistic regression	68%	90.8%	73.1%	77.1%	75.7%	89.5%
Ordinal logistic regression	53.7%	90%	67.8%	78.1%	57.2%	87.5%
Linear discriminant Analysis	59.3%	90.7%	72.5%	77.9%	65.8%	87.8%
Quadratic discriminant analysis	69.9%	90.4%	76.8%	71.7%	69.94%	86.8%

Table 3: Classification accuracy of predictions based on biomarker concentrations for healthy, high-risk and mixed populations. For example, the Quadratic discriminant analysis model can predict the correct time since administration category in the healthy population with 69.9% accuracy.

Figure 8 shows how the probability of liver injury changes over time for both healthy and high-risk populations, for doses between 0-1000 mg/kg. The model-derived probabilities are dependent only on HMGB1 concentration (as predicted by our previous logistic regression model [20]). A threshold probability of 0.5 is used to determine the likelihood of DILI. Any observation within the white contour boundary is therefore predicted to be representative of probable liver injury.

A



B

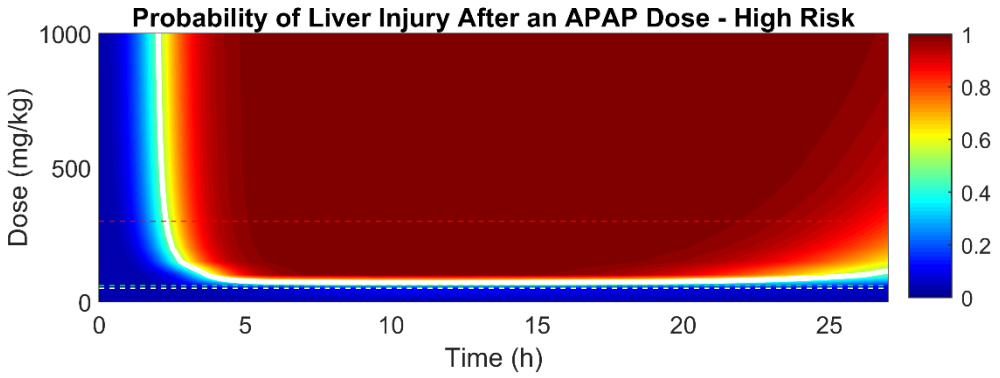


Figure 8. **Proposed framework for predicting probability of liver injury dependent upon dose, time and HMGB1 concentration.** The plotted frameworks are representative of both healthy (A) and high-risk (B) populations. In each, the white contour indicates the threshold of 0.5 probability of liver injury; the red dashed-line represents the currently used APAP dose for toxicity studies in mice; the white dashed-line represents toxic dose as proposed by our model; and the green dashed-line indicates currently considered therapeutic dose for mice. In the high-risk population, the toxic dose proposed by our model and the therapeutic dose are identical.

For the healthy population, the time-frame for biomarker concentrations representing those of probable liver injury is around 2-18 hours. The dose threshold at which toxicity is predicted to be likely to occur is approximately 350 mg/kg. Note that this prediction is only slightly above the currently used toxic dose (300 mg/kg) [44]. The framework for a healthy population suggests that there is almost 100% chance of liver injury when the dose is only slightly higher than this amount.

For the fasted population, however, at the currently recommended therapeutic APAP dose (60 mg/kg), there is approximately 50% chance of liver injury progression. For any dose above the known therapeutic threshold, liver injury progression is predicted to be highly probable (above 60%); beyond approximately 2.5 h post-dose, HMGB1 concentrations remain indicative of highly probable liver injury for the whole time-course. These results are summarised in Table 4.

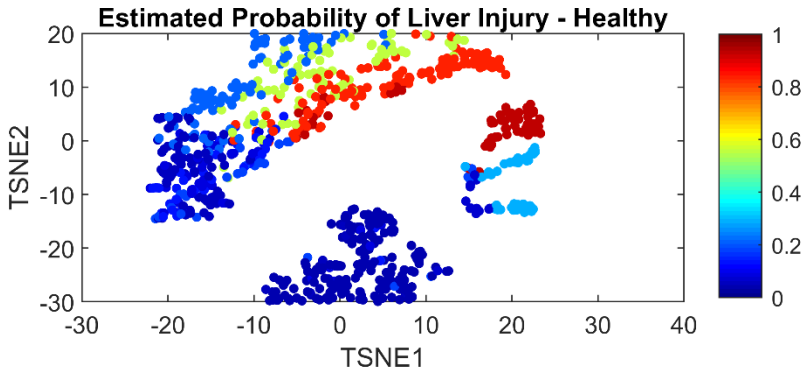
Population	Dose likely to induce liver injury	Time frame liver injury likely to occur
Healthy	>350 mg/kg	2-18 h
High-risk	>60 mg/kg	>2.5 h

Table 4: Critical dose and time ranges indicative of potential liver injury in both healthy and high-risk populations.

3.6 Visualising the probability of liver injury following an APAP dose

For each observation within the *in-silico* population, simulated HMGB1 concentrations, based on our findings in [20], are used within the logistic regression model to determine the corresponding probability of liver injury. Each *in-silico* observation is then visualised and discriminated by their resultant liver injury probability using the TSNE method (Figure 9). Relatively safe/unsafe observations are identifiable in both healthy and high-risk populations. For the healthy population (Figure 9A) most of the observations have less than around 35% chance of DILI progression. The small group of observations representing at-risk individuals can be identified at the top, right-hand side of the image, portrayed by red markers. In stark contrast, most of the *in-silico* observations have almost 100% chance of liver injury progression in the high-risk population, as can be seen by the dominance of red markers in Figure 9B. The observations corresponding to lower injury potential are indicated in the central and the left-hand side of the plot.

A



B

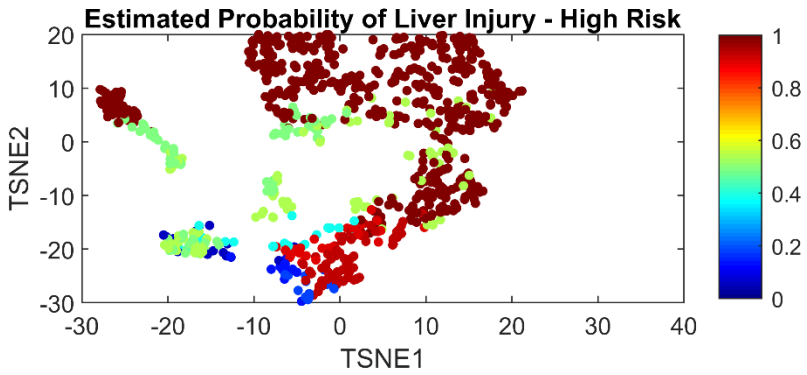


Figure 9: 2-dimensional TSNE visualisation of *in-silico* observations with respect to estimated probability of liver injury. Separated liver injury probabilities are visualised for a range of *in-silico* generated inputs for both healthy (A) and high-risk (B) populations. DILI probabilities (0-1) for each individual observation are indicated by the colour bar.

4 Discussion

The current clinical APAP framework is known to be inaccurate, mainly due to a dependency on known time of ingestion, but also due to inaccurate estimates of initial dose based on biomarkers currently criticised for being insensitive [25]. We previously reported a promising computational framework for predicting the probability of liver injury based on novel biomarker concentrations [20]. However, the identifiability analysis performed within this study highlights parameter unidentifiabilities within the model that required addressing prior to any further clinical application. Results of the analysis show where the original model required structural changes and also where more data was required in order to increase confidence in

predictions. Therefore, we have re-structured the original model where necessary and re-parameterised against additional data (from both fed and fasted mice). Identifiability analysis of the newly defined model structure found that only 7% of parameters are practically unidentifiable (compared with 24% previously).

As well as improving confidence in the mathematical representation of the system, we have here provided a model that is more representative of the clinical environment by including a wider range of APAP doses. Furthermore, the new treatment framework can now be adjusted for individuals considered to be at high risk of APAP-induced liver injury [3]. To account for the fact that many overdose patients have an increased underlying susceptibility to APAP-induced liver injury, we optimise our model against both fed and fasted mouse data. Factors that result in an APAP overdose patient being deemed “high-risk” such as alcohol consumption or malnutrition are known to deplete intracellular GSH stores, which subsequently increases their risk of potential toxicity [45]. Fasted animals have also been shown to have depleted GSH stores [46] and hence they are considered to be representative of high-risk individuals in our framework. When training the biomarker models against multiple datasets and comparing to the original model [20], which is calibrated against only 530 mg/kg APAP dose time-course data, there are noticeable adjustments in the parameter values, particularly for the GSH model. Many of these changes are intuitive and may be representative of changes in mechanisms due to fasting. Additionally, changes may be attributable to the fact that the model can now better account for a larger variety of dosing scenarios. The significant changes in parameter values highlight the importance of optimising against a range of both therapeutic and toxic datasets and both healthy and unhealthy populations.

The model we propose in this paper is more biologically relevant than our previous model [20]. Originally, we assume that full and fragmented K18 have similar mechanisms and act as independent biomarkers. However, full and fragmented K18 are known to be effectively necrotic/apoptotic versions of the same biomarker [47]. Incorporating this into our model

framework we confirm that necrosis is the pre-dominant form of cell death in mice APAP overdose cases [48]. An identifiability analysis on the new model structure identifies an increased confidence in parameter estimates for the GSH, HMGB1, full and fragmented K18 components of the model structure. There are, however, some unidentifiabilities remaining, particularly within the HMGB1 and fragmented K18 components of the model, indicating that additional data and model development is still required in order to have full confidence in the uniqueness of chosen mechanistic parameter values.

We provide a proof of-concept framework which can be used to make toxicity predictions for either a healthy, high-risk or mixed population. In practice, once a dose/time is determined using the multiple linear regression statistical model, probability of liver injury can be estimated based on the HMGB1-dependent logistic regression model, driven by PK simulations with the previously identified dose/time values. The range of doses for which the high-risk population is likely to suffer liver injury (based on predictions made within this *in-silico* framework) is extensive. Prediction accuracy of time since APAP administration and initial APAP dose is similar for the healthy and mixed populations, with a 3-4 h error for time since administration and a 65-75 mg/kg error for initial dose. The initial dose is much harder to predict in the high-risk population (184.1 mg/kg error). This result is unsurprising since a much larger range of doses will have a toxic effect if the liver is already impaired. Both Figure 8 and Table 4 show that in a high-risk population, any dose above approximately 60 mg/kg is highly likely to induce liver injury and therefore determining the exact dose in this scenario is difficult. Time since administration, however, could be predicted more accurately than in a healthy/mixed population (2.118 h error). It has already been found that amending treatment thresholds to account for high-risk individuals can better protect those with greater liver injury susceptibility [3]. Table 4 shows that liver injury in a healthy individual is predicted to occur within 2-18 hours following an APAP dose. For a high-risk individual however, this likelihood is predicted to continue for the whole time-course investigated (24 hours). Results from this study further

643 endorse the idea that there are likely very different outcomes with respect to liver injury
644 potential for healthy and high-risk individuals. More-informed decisions can therefore be made
645 regarding optimal treatment if clinicians can identify those who are more susceptible to
646 overdose. This would significantly improve patient outcomes while reducing the cost and
647 burden of unnecessary antidote treatment.

648 The model was recalibrated to improve confidence in our approach and ensure that predictions
649 are more clinically applicable prior to developing a clinical extension of the framework. The
650 model has been calibrated against mouse data due to the quantity and quality of the toxicity
651 data available, necessary for the rigorous development of a mathematical framework. Whilst
652 metabolic similarities do exist between rodent and human (for example CYP2E1 expression),
653 expression levels of enzymes such as CYP1A2 are known to vary across species [49],
654 meaning direct translation of rodent-based models to the human case is rarely feasible and
655 additional work may be required. Animal models account for the complex, physiological
656 interplay within organisms, and therefore are often investigated during pharmacological
657 research prior to application in humans. Although clinical APAP data is available, it is often
658 sparse, erroneous and inherently highly variable. A Population-Pharmacokinetics (Pop-PK)
659 approach could potentially provide insight into the stochasticity of the errors involved, but
660 initially we have focused on the mouse framework as a convenient testing and development
661 toolkit to determine efficacy in a well-controlled environment. The promising results of this
662 study now provide confidence in the feasibility of translating our approaches to the human
663 clinical case. Identifiability analysis has provided insight into areas within our model structure
664 that remain uncertain, and therefore require additional consideration and improvement in order
665 to enhance confidence in translated predictions. Regarding translation, all biomarkers used in
666 the study can be measured in both animals and humans via the same methodologies.
667 Inevitably, there are apparent differences between humans and mice such as the expression
668 of metabolising enzymes, the mass APAP dose required to induce toxicity and the kinetics of

669 the biomarker profiles [44, 45]. However, the essential mechanistic processes of APAP
670 metabolism, toxicity and action of the antidote are directly comparable. Our model accounts
671 for a dose range 0-1000 mg/kg in mouse and this can be amended as a parameter within our
672 model structure to be applicable to the clinical situation. In terms of the biomarkers used, they
673 have now been measured in a time-dependent manner in multiple human studies [33, 41]. In
674 its current form, our framework accounts for varied dose ranges and provides promise for
675 clinical use in determining initial dose, time since ingestion and estimated probability of liver
676 injury for both healthy and high-risk individuals. Further model development, integration of
677 additional experimental data and translation to the clinical environment is now required to
678 significantly advance this research to provide a state-of-the-art alternative to the methods
679 currently used and considerably improve individualised treatment of APAP overdose.

680 **Conflict of interest**

681 CLM is now an employee of AstraZeneca PLC.

682 **Acknowledgements**

683 CLM acknowledges funding support from the Faculty of Engineering and Technology,
684 Liverpool John Moores University. JL and SDW acknowledge funding support from the
685 Liverpool Centre for Mathematics in Healthcare (EPSRC grant: EP/N014499/1). JL is
686 supported by an MRC Skills Development Fellowship (MR/S019332/1). CLM acknowledges
687 Professor Jens Timmer's group from The Freiburg Institute for Advanced Studies (FRIAS) for
688 an introduction to the technique of identifiability analysis.

689 **Author Contributions**

690 CLM contributed to the mathematical modelling and statistical analysis and wrote the
691 manuscript; JL contributed to the mathematical modelling; ST and IJ contributed to the

statistical analysis; SDW contributed to the mathematical modelling and designed the research. All authors read and approved the final manuscript.

695 **References**

696 [1] A. Chiew, J. Fountain, A. Graudins, K. I. Geoffrey, D. Reith, and N. A. Buckley,
697 “Guidelines for the management of paracetamol poisoning in Australia and New
698 Zealand,” *Med. J. Aust.*, 2015.

699 [2] M. Yan, Y. Huo, S. Yin, and H. Hu, “Mechanisms of acetaminophen-induced liver injury
700 and its implications for therapeutic interventions,” *Redox Biol.*, vol. 17, no. March, pp.
701 274–283, 2018.

702 [3] E. Yoon, A. Babar, M. Choudhary, M. Kutner, and N. Pyrsopoulos, “Acetaminophen-
703 Induced Hepatotoxicity: a Comprehensive Update,” *J. Clin. Transl. Hepatol.*, vol. 4, no.
704 2, pp. 131–142, 2016.

705 [4] N. Buckley and M. Eddleston, “Paracetamol (acetaminophen) poisoning,” *Clin. Evid.*
706 *(Online).*, no. March, pp. 1–14, 2007.

707 [5] ONS, “Deaths related to drug poisoning in England and Wales: 2016 registrations,”
708 *statistical bulletin*, 2016. [Online]. Available:
709 https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/de
710 aths/bulletins/deathsrelatedtodrugpoisoninginenglandandwales/2016registrations.
711 [Accessed: 11-Sep-2018].

712 [6] W. M. Lee, “Acetaminophen and the U.S. acute liver failure study group: Lowering the
713 risks of hepatic failure,” *Hepatology*, vol. 40, no. 1, pp. 6–9, 2004.

714 [7] B. Rumack, R. Peterson, G. Koch, and I. Amara, “Acetaminophen overdose. 662 cases
715 with evaluation of oral acetylcysteine treatment.,” *Arch. Int. Med.*, vol. 141, no. 3, pp.
716 380–385, 1981.

717 [8] A. Algren, “Review of N-Acetylcysteine for The Treatment of Acetaminophen

718 (Paracetamol) Toxicity in Pediatrics." Expert Commitee on The Selection and Use of
719 Essential Medicines, 2008.

720 [9] D. N. Bateman *et al.*, "Effect of the UK's revised paracetamol poisoning management
721 guidelines on admissions, adverse reactions and costs of treatment," *Br. J. Clin.*
722 *Pharmacol.*, vol. 78, no. 3, pp. 610–618, 2014.

723 [10] L. L. Mazaleuskaya, K. Sangkuhl, C. F. Thorn, G. A. Fitzgerald, R. B. Altman, and T. E.
724 Klein, "PharmGKB summary: Pathways of acetaminophen metabolism at the
725 therapeutic versus toxic doses," *Pharmacogenet. Genomics*, vol. 25, no. 8, pp. 416–
726 426, 2015.

727 [11] R. Ferner, W. J. Dear, and N. Bateman, "Management of paracetamol poisoning," *BMJ*,
728 vol. 342, pp. 1–19, 2011.

729 [12] B. H. Rumack, "Acetaminophen overdose," *Am.J.Med.*, vol. 75, no. 0002–9343, pp.
730 104–112, 1983.

731 [13] Beat, "Eating disorder statistics," 2018. [Online]. Available:
732 https://www.beateatingdisorders.org.uk/media-centre/eating-disorder-statistics.
733 [Accessed: 11-Sep-2018].

734 [14] W. M. Lee, "Acetaminophen (APAP) hepatotoxicity—Isn't it time for APAP to go away?,"
735 *J. Hepatol.*, vol. 67, no. 6, pp. 1324–1331, 2017.

736 [15] NHS, "Statistics on alcohol, England 2017," 2017. [Online]. Available:
737 http://webarchive.nationalarchives.gov.uk/20180328130416/http://digital.nhs.uk/catalo
738 gue/PUB23940. [Accessed: 12-Sep-2018].

739 [16] D. Callahan, "Managed Care and the Goals of Medicine," *Healthplan*, vol. 38, no. 5, p.
740 9,17, 1997.

741 [17] NC3Rs, “The 3Rs,” 2017. [Online]. Available: https://www.nc3rs.org.uk/the-3rs.
742 [Accessed: 28-Apr-2017].

743 [18] P. Bloomingdale *et al.*, “Quantitative systems toxicology,” *Curr. Opin. Toxicol.*, vol. 4,
744 pp. 79–87, 2017.

745 [19] J. Vanlier, *Uncertainty Analysis in Systems Biology*. Eindhoven: Technische Universiteit
746 Eindhoven, 2014.

747 [20] C. L. Mason, J. Leedale, S. Tasoulis, I. Jarman, D. J. Antoine, and S. D. Webb,
748 “Systems Toxicology Approach to Identifying Paracetamol Overdose,” *CPT*
749 *Pharmacometrics Syst. Pharmacol.*, vol. 7, no. 6, pp. 394–403, 2018.

750 [21] B. A. Howell *et al.*, “In vitro to in vivo extrapolation and species response comparisons
751 for drug-induced liver injury (DILI) using DILIsymTM: A mechanistic, mathematical
752 model of DILI,” *J. Pharmacokinet. Pharmacodyn.*, vol. 39, no. 5, pp. 527–541, 2012.

753 [22] D. Reith, N. J. Medlicott, R. Kumara De Silva, L. Yang, J. Hickling, and M. Zacharias,
754 “Simultaneous modelling of the michaelis-menten kinetics of paracetamol sulphation
755 and glucuronidation,” *Clin. Exp. Pharmacol. Physiol.*, vol. 36, no. 1, pp. 35–42, 2009.

756 [23] J. G. Diaz Ochoa, J. Bucher, A. R. R. Péry, J. M. Zaldivar Comenges, J. Niklas, and K.
757 Mauch, “A multi-scale modeling framework for individualized, spatiotemporal prediction
758 of drug effects and toxicological risk,” *Front. Pharmacol.*, vol. 3, no. 204, pp. 1–11, 2013.

759 [24] T. J. Zurlinden, K. Heard, and B. Reisfeld, “A novel approach for estimating ingested
760 dose associated with paracetamol overdose,” *Br. J. Clin. Pharmacol.*, vol. 81, no. 4, pp.
761 634–645, 2016.

762 [25] J. W. Dear *et al.*, “Risk stratification after paracetamol overdose using mechanistic
763 biomarkers : results from two prospective cohort studies,” *Lancet Gastroenterol.*

764 *Hepatol.*, vol. 1253, no. 17, pp. 1–10, 2017.

765 [26] C. H. Remien, F. R. Adler, L. Waddoups, T. D. Box, and N. L. Sussman, “Mathematical
766 modeling of liver injury and dysfunction after acetaminophen overdose: Early
767 discrimination between survival and death,” *Hepatology*, vol. 56, no. 2, pp. 727–734,
768 2012.

769 [27] A. Navid, D. M. Ng, B. J. Stewart, S. E. Wong, and F. C. Lightstone, “Quantitative In
770 Silico analysis of transient metabolism of acetaminophen and associated causes of
771 hepatotoxicity in humans,” *Silico Pharmacol.*, vol. 1, no. 1, p. 14, 2013.

772 [28] T. J. Zurlinden and B. Reisfeld, “Characterizing the effects of race/ethnicity on
773 Acetaminophen pharmacokinetics using physiologically based pharmacokinetic
774 modelling,” *Eur. J. Drug Metab. Pharmacokinet.*, vol. 42, no. 1, pp. 143–153, 2017.

775 [29] M. Coen *et al.*, “Integrated application of transcriptomics and metabonomics yields new
776 insight into the toxicity due to paracetamol in the mouse,” *J. Pharm. Biomed. Anal.*, vol.
777 35, no. 1, pp. 93–105, 2004.

778 [30] D. J. Antoine *et al.*, “High-mobility group box-1 protein and keratin-18, circulating serum
779 proteins informative of acetaminophen-induced necrosis and apoptosis in vivo,” *Toxicol.
780 Sci.*, vol. 112, no. 2, pp. 521–531, 2009.

781 [31] D. J. Antoine, D. P. Williams, A. Kipar, H. Laverty, and B. K. Park, “Diet Restriction
782 Inhibits Apoptosis and HMGB1 Oxidation and Promotes Inflammatory Cell Recruitment
783 during Acetaminophen Hepatotoxicity,” vol. 16, no. December, pp. 479–490, 2010.

784 [32] Mathworks, “Matlab.” The MathWorks Inc., Natick, Massachusetts, 2016.

785 [33] M. H. Wright, “Nelder, Mead, and the Other Simplex Method,” 2010.

786 [34] A. Raue *et al.*, “Structural and practical identifiability analysis of partially observed

787 dynamical models by exploiting the profile likelihood," *Bioinformatics*, vol. 25, no. 15,
788 pp. 1923–1929, 2009.

789 [35] A. Raue *et al.*, "Identifiability and observability analysis for experimental design in
790 nonlinear dynamical models," *Chaos*, vol. 20, 2010.

791 [36] A. Sharma and W. J. Jusko, "Characteristics of indirect pharmacodynamic models and
792 applications to clinical drug responses," *Br. J. Clin. Pharmacol.*, vol. 45, no. 3, pp. 229–
793 239, 1998.

794 [37] D. M. Longo *et al.*, "Refining Liver Safety Risk Assessment : Application of Mechanistic
795 Modeling and Serum Biomarkers to Cimaglermin Alfa (GGF2) Clinical Trials," vol. 102,
796 no. 6, pp. 961–969, 2017.

797 [38] D. G. N. Craig, C. M. Bates, J. S. Davidson, K. G. Martin, P. C. Hayes, and K. J.
798 Simpson, "Overdose pattern and outcome in paracetamol- induced acute severe
799 hepatotoxicity," 2011.

800 [39] C. Saranya and M. Guimarães, "A Study on Normalization Techniques for Privacy
801 Preserving Data Mining," *Int. J. Eng. Technol.*, vol. 5, pp. 2701–2704, 2015.

802 [40] J. Franc, "Robust regression - Robust estimation of regression coefficients in linear
803 regression model when orthogonality condition is breaking." pp. 1–45, 2017.

804 [41] W. Ambrosius, "Multiple Linear Regression," in *Topics in Biostatistics*, 2007, pp. 165–
805 187.

806 [42] L. Van Der Maaten, "Accelerating t-SNE using tree-based algorithms," vol. 15, pp.
807 3221–3245, 2014.

808 [43] M. Brannick, "Logistic Regression," 2016. [Online]. Available:
809 http://faculty.cas.usf.edu/mbrannick/regression/Logistic.html. [Accessed: 28-Apr-

2017].

[44] J. A. Hinson, D. W. Roberts, and L. P. James, “Mechanisms of Acetaminophen-Induced Liver Necrosis,” *Pharmacology*, vol. 196, no. 196, pp. 1–34, 2010.

[45] D. Davidson, “Association of acetaminophen hepatotoxicity with fasting and ethanol use.,” *J. Emerg. Med.*, pp. 737–738, 1995.

[46] R. M. Walker, T. E. Massey, T. F. McElligott, and W. J. Racz, “Acetaminophen toxicity in fed and fasted mice.,” *Can. J. Physiol. Pharmacol.*, vol. 60, no. 3, pp. 399–404, 1982.

[47] R. D. Beger *et al.*, “Translational biomarkers of acetaminophen-induced acute liver injury,” *Arch. Toxicol.*, vol. 89, no. 9, pp. 1497–1522, 2015.

[48] J. S. Gujral, T. R. Knight, A. Farhood, M. L. Bajt, and H. Jaeschke, “Mode of cell death after acetaminophen overdose in mice: Apoptosis or oncotic necrosis?,” *Toxicol. Sci.*, vol. 67, no. 2, pp. 322–328, 2002.

[49] M. Martignoni, G. M. M. Groothuis, and R. De Kanter, “Species differences between mouse , rat , dog , monkey and human CYP-mediated drug metabolism , inhibition and induction,” *Expert Opin. Drug Metab. Toxicol.*, vol. 2, no. 6, pp. 875–894, 2006.

[50] X. Ge *et al.*, “High Mobility Group Box-1 (HMGB1) Participates in the Pathogenesis of Alcoholic Liver Disease (ALD) *,” vol. 289, no. 33, pp. 22672–22691, 2014.

[51] N. Brillant *et al.*, “Dynamic and accurate assessment of acetaminophen-induced hepatotoxicity by integrated photoacoustic imaging and mechanistic biomarkers in vivo,” *Toxicol. Appl. Pharmacol.*, vol. 332, pp. 64–74, 2017.